The Many Faces of Natural Language Searching

Natalie A. Schoch and Winifred Sewell
College of Library and Information Services, University of Maryland, College Park, MD

Recently a number of online services have begun to market natural language searching aimed at end users. Although the systems vary, all accept queries in the searcher's own language and rank the results with the "best" reference first. To do so, each uses its own matching algorithm for weighting and determining frequency of the query words in the database. Unlike Boolean systems, the natural language search engines do not require that all concepts in a question be present in every reference. Some natural language systems stem query terms and add synonyms to the search.

Because the systems differ, the authors wished to learn what they have in common and what to tell end users about searching them effectively. To do so, we chose the four systems known to search MEDLINE in natural language: FreeStyle from Lexis-Nexis (FS), Knowledge Finder from Aries (KF), Physicians' Online (POL), and Target on Dialog from Knight-Ridder (TA). All are online systems except KF, which is available only on CD-ROM. We chose 36 topics to provide a broad scope in subject areas within the health sciences as well as variety in levels of specificity and expected number of matching references. Exhaustive searches for each topic were performed on MEDLINE directly from the National Library of Medicine (NLM). On the same day during January or February 1995, each search was updated on the NLM database and run in its simplest form on all four natural language systems. The terms used were identical for the four systems, except that we complied with the request of FS and TA that quotation marks be placed around words that should appear together. The first 25 ranked references from each search were selected for analysis. Although all the systems provide Boolean searching as well, we used only natural language.

A complete list of the topics searched and quotation marks used for FS and TA will be included in another paper that will summarize and compare the results and postulate reasons for better retrieval with some topics. The present paper provides examples of advantages and hazards of natural language searching for the topics we used.

All four systems agreed on the best reference for only one topic, hearing or hearing disorders in premature infants. In three other instances, three systems (FS, KF, and TA) chose the same first reference. POL ranked it second twice, but for the third topic, did not include it at all. The queries covered cardiac effects of selenium, PUVA therapy in dermatitis, and aloe in dermatitis. A more likely scenario was for a reference to appear in only one or two of the four natural language searches. In the premature infant hearing question, for instance, only six references were agreed upon by two or more search engines. When two or three chose a citation, they might rank it quite differently.

Two systems (FS and TA) provide some insight into their ranking methods by showing weights and occurrences (FS) and by providing percentage ranks and frequency counts (TA). In a question on the use of orthodontic appliances in temporomandibular joint syndrome, TA's first six references matched both concepts, having statistical relevance ranging from 99% to 31%. The remainder had high frequencies of occurrence for one or the other of the two concepts, but no co-occurrence of both. In a search for effects of antibiotics in cattle on their meat, FS did not find any references that matched the question, presumably because it gave meat a rank of 14, cattle, 3, and antibiotics, 4. Meat occurred in all 25 of the ranked references, cattle in 24, and antibiotics in only the 13th, on unspecified raw meat for racing greyhounds. The four searches found a total of 12 unique references with all concepts matching. Only four of these had appeared in the NLM search, but it retrieved 11 other good references through the explosion of the term antibiotics.

Two of the systems (KF and POL) used stemming liberally and also added synonyms. TA requested us to provide synonyms, but -- mimicking end users we know -- we did not do so, electing to use exactly the same question in all systems. The synonyms were helpful for calcification of artificial heart valves, in which KF found papers on heart valve prostheses missed by others. Disadvantages of stemming are shown by KF's retrieval of posterior urethral valve (PUV) imaging for PUVA therapy in dermatitis.

Further examples will be available with the poster as well as a list of all questions and details of examples.

0195-4210/95/$5.00